

# DL-driven pattern recognition workflow in Pandora

Andy Chappell

01/03/2021

FD Sim/Reco Meeting

# Overview

The logo for Warwick University, featuring a stylized blue zigzag line above the word "WARWICK" in a blue, sans-serif font.

- Testing application of hit-level track/shower ID to 2D clustering
- Split reconstruction into track and shower streams
- Cheat different aspects of split reconstruction
- Tuning clustering algorithms in shower stream
- New 2D clustering algorithm

# Updated pattern recognition structure



- Create initial 2D clusters
- Assign clusters to either track stream, or shower stream
- Run 2D clustering algorithms on each stream independently

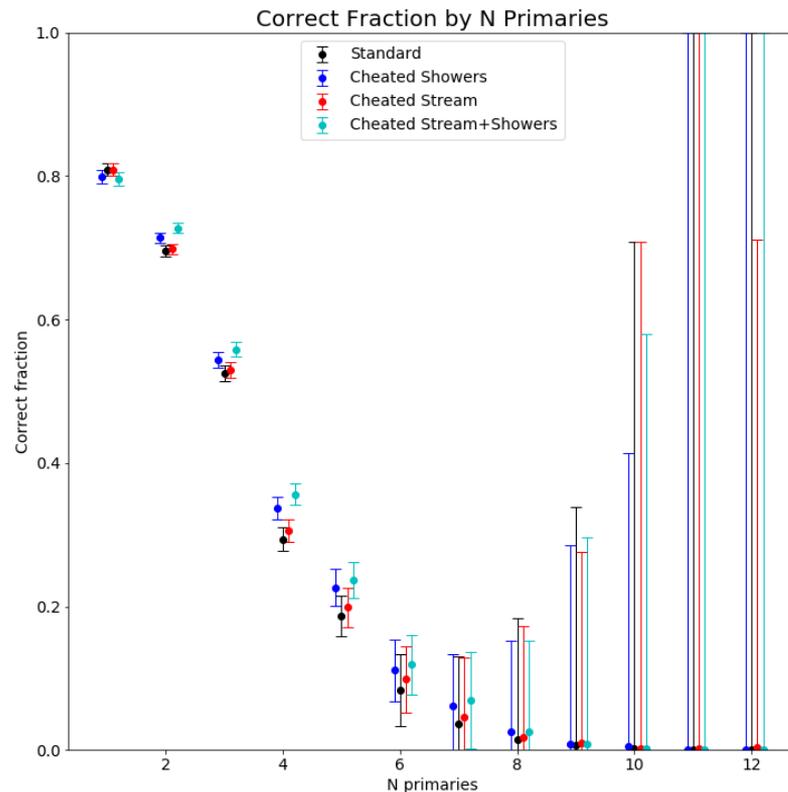
Standard	Cheating		
	Stream	Shower	Both
No track/shower streaming	Cheat assignment of clusters to each stream	Network assignment of clusters to each stream	Cheat assignment of clusters to each stream
	Standard 2D clustering within each stream	Standard 2D clustering for track stream	Standard 2D clustering for track stream
		Cheated 2D clustering for shower stream	Cheated 2D clustering for shower stream
Standard reconstruction post 2D clustering	Standard reconstruction post 2D clustering	Standard reconstruction post 2D clustering	Standard reconstruction post 2D clustering

# End of reco correct fraction

WARWICK

- 100K events split evenly between numu and nue DUNE FD MCC11 1x2x6
- An event is deemed correct if
  - There is exactly one reconstructed particle for each MC particle
  - And those reconstructed particles must each be at least 50% pure
- Modest improvement in correct fraction for cheated cases\*
- Similar performance when cheating showers for cheated and network stream assignment

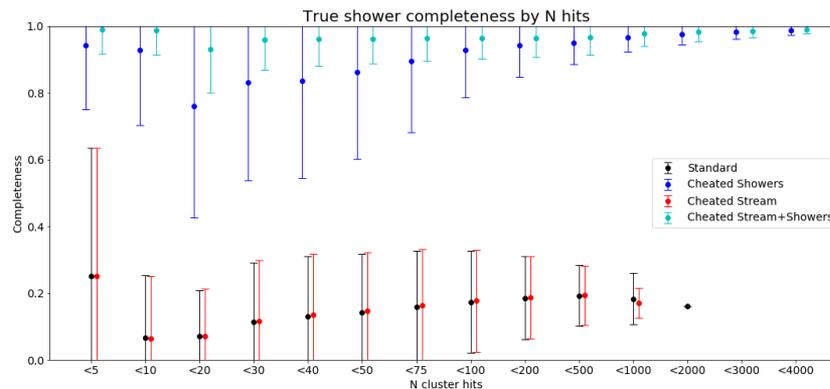
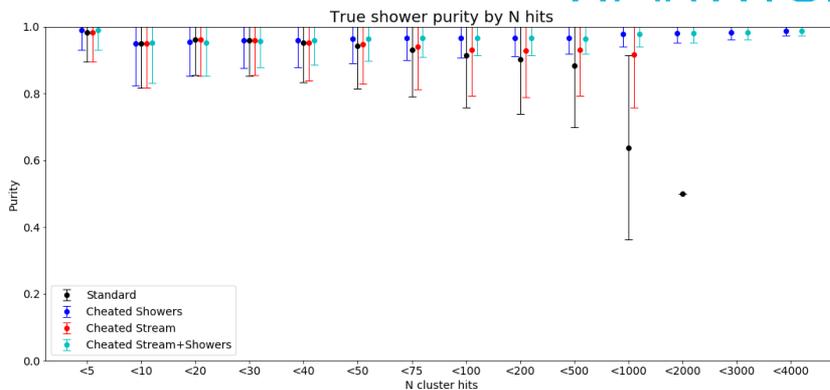
\* Cheating only 2D clustering, yet to test cheating PFO creation in the shower stream



# Cluster purity and completeness

WARWICK

- Assess end of 2D clustering performance
- Clusters in the standard reconstruction are very pure
- Clusters in the three cheated cases are similarly pure
- Clusters in the standard reconstruction have low completeness
- Clusters in the shower reconstruction cheating cases have high completeness
  - Network stream assignment is good, but clearly still some scope for improvement

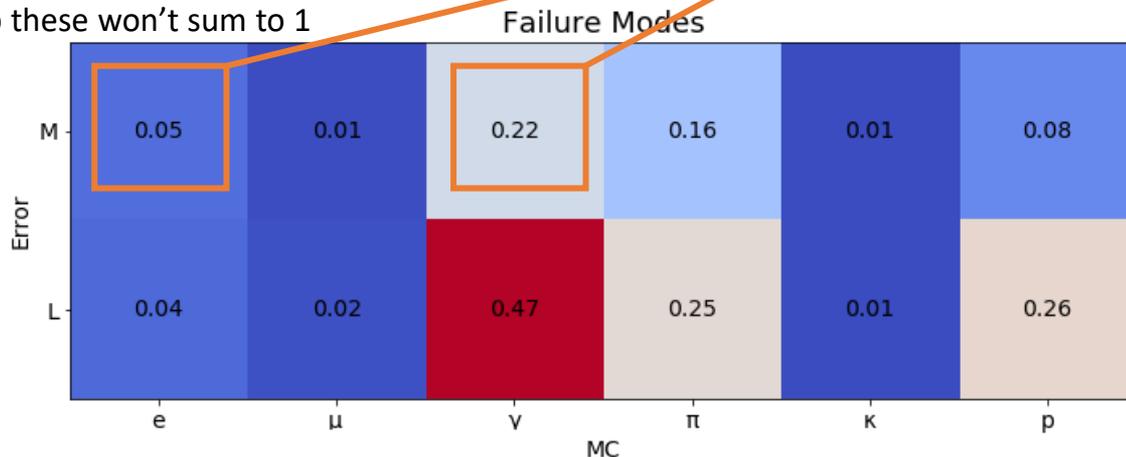


# Standard reconstruction failure modes

WARWICK

- Useful to identify the nature of the errors in pattern recognition
- Identify two broad categories of error
  - M: Incorrect event due to MC particle being split amongst several (high purity) reconstructed particles or having an insufficiently pure reconstructed particle
  - L: Incorrect event due to an MC particle having no matching reconstructed particle
- Determine the fraction of **events** that contain each failure mode
  - These aren't exclusive failure modes, so these won't sum to 1
- Half of all incorrectly reconstructed events have at least one unreconstructed photon
- A quarter fail to achieve a high purity one-to-one match to at least one photon

Cheating is fixing these cases

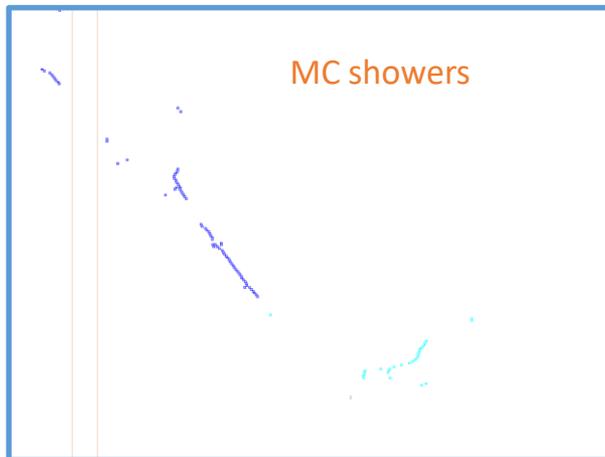
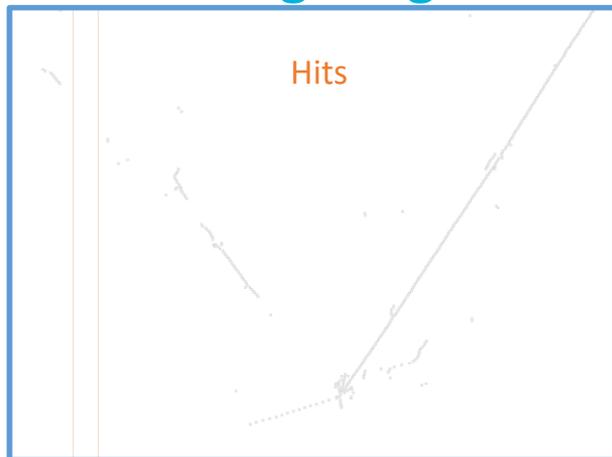


# Tuning streamed clustering



- Starting to look for improvements in 2D clustering of shower-like clusters
- Take advantage of the fact that we 'know' we're dealing with shower-like topologies in one of the streams
- Phase I – tune the existing (track-focused) reconstruction algorithms
  - Longitudinal association and extension algorithms can be more generous in merging clusters
  - We probably don't want to split kinked clusters as readily
- Phase II – introduce new algorithms to identify the topological features associated with showers

## Tuning longitudinal association criteria



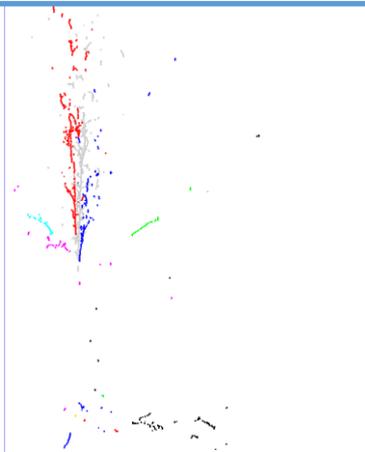
# Limits of tuning existing algorithms

WARWICK

Hits



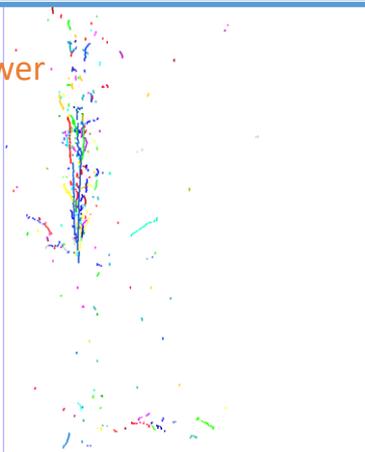
MC showers



Standard clusters



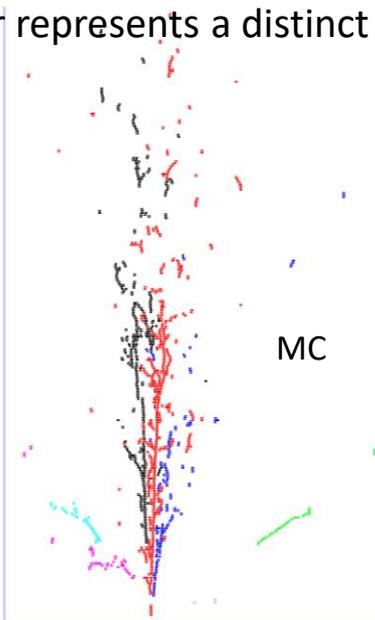
Streamed shower clusters



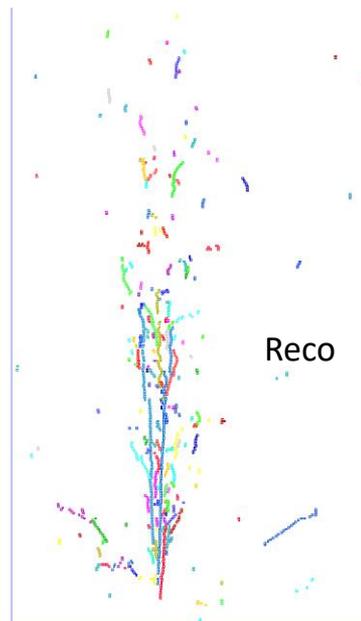
2D clustering algorithms are track-centric, they aren't designed to cope with topologies like this

## New clustering algorithm in development

- Focusing on shower induced by  $\sim 6$  GeV primary pion in event from previous slide
- Each cluster represents a distinct leading shower

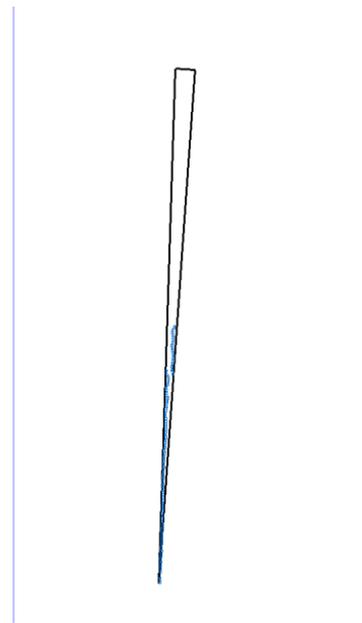
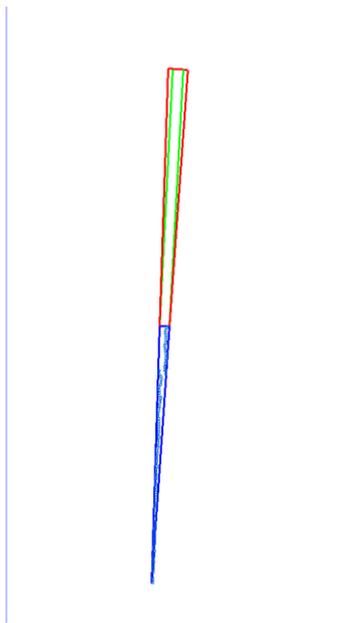
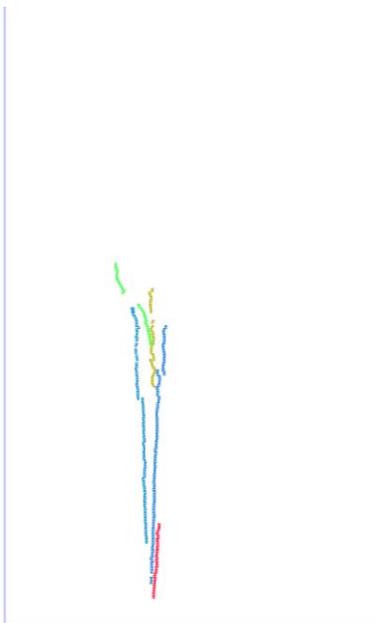


- Input clusters to shower stream



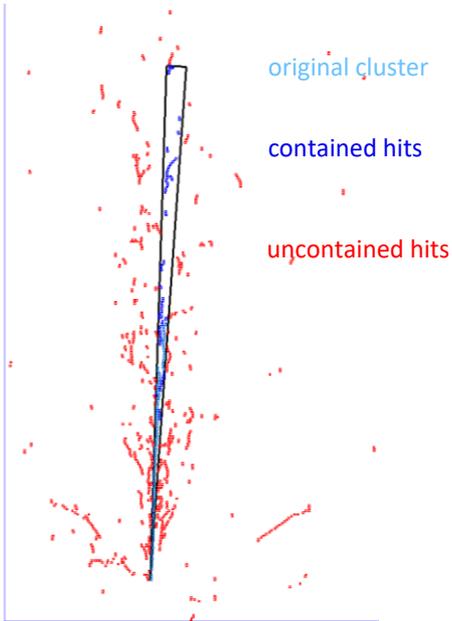
## PCA-based cluster growing

- Focus on clusters with > 30 hits as seeds
- Define bounding regions from PCA
- Focus on extended bounds



# PCA-based cluster growing

- Check bounds hit intersection



- From this seed...



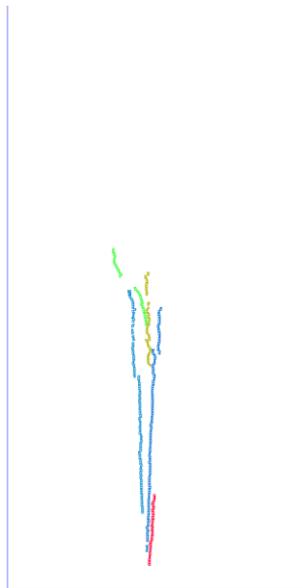
- We get this



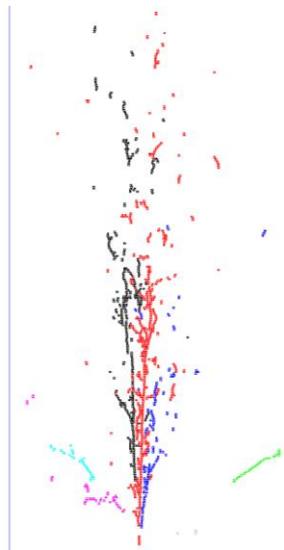
- Apply a second pass with broader transverse bounding region and no longitudinal projection...

# PCA-based cluster growing

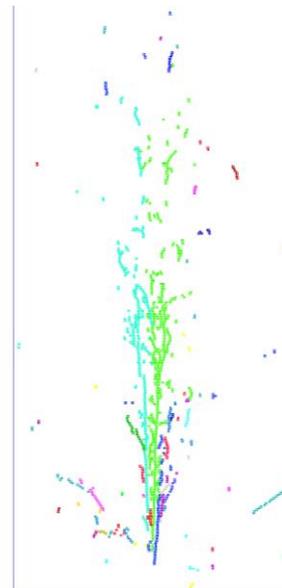
- Input seed clusters



- MC



- Output clusters



## Summary



- Scope to increase completeness of 2D shower clusters with track/shower separation into independent streams
- Increased 2D completeness can reduce event reconstruction failures caused matching multiple PFOs to a single MC particle
- Some scope to tune existing track-centric algorithms to operate in a shower environment
- Developing new algorithms to fully exploit the separation